Encrypting Columns in an Rdb Database

Jeffrey S. Jalbert

JCC Consulting, Inc.

Abstract

Reports of lost data and data theft have reached a crescendo in the past few months. Rules for managing data, especially those containing enough information to allow identify theft are rising and, in particular, credit suppliers such as VISA and Master Card require enterprises storing this data to certify that personal identifying information is secure.

This presentation will discuss our methodology for implementing column-level encryption within an Rdb database and encrypting VMS backups for one JCC customer.

The research reported was accomplished by JCC for one customer and examples reference that customer's data.

Business Issues

- Identity theft has become a big-time issue
  - A Google search yielded 17.2 million hits on the subject
  - Specific instances “close by” (customer is in Ohio)
    - State of Ohio - the recent theft of a state data storage device.
    - Ohio University – one of the compromised servers, which held Social Security numbers belonging to 137,000 people, was penetrated by U.S. and overseas-based hackers for at least a year and possibly much longer
  - Many more examples, almost weekly

Headlines

- MSNBC - America's veterans were sent scrambling for their credit reports Monday, as the Veteran's administration announced nearly all of them — and some of their family members — were at heightened risk for identity theft.
- According to the Federal Trade Commission (FTC), identity theft cost consumers and businesses $52.6 billion in 2004 alone

Business Requirement

- Any enterprise accepting payments via credit card must encrypt identifying information such as SSN. (Payment Card Industry (PCI) Compliance)
  - Google search on PCI compliance yielded 26,700 hits!
- Public Relations – You can trust us, we encrypt identifying data

Logical Issues

- Cryptography (theory of encryption) is an obscure subject
  - Highly mathematical, involving theory of prime numbers
  - Cryptography today might be summed up as the study of techniques and applications that depend on the existence of difficult problems

Encryption

- *Encryption* is the transformation of data into a form that is as close to impossible as possible to read without appropriate (secret) knowledge (a key)

Some Sources

- RSA Laboratories has a very informative FAQ on Cryptography
  - http://www.nordugrid.org/documents/rsalabs_faq41.pdf
- G. Brassard, *Modern Cryptology*, Springer-Verlag, 1988
- R.L. Rivest, Cryptography, *Handbook of Theoretical Computer Science, volume A*

Cryptography Standards

- Standards allow different vendors to interact together
- Provide a basis for testing the quality of algorithms
  - Complexity implies potential error
- Many standards organizations (ISO, ANSI, IEEE, NIST, IETF)

Cryptography Standards

- IBM & US Department of Commerce (NIST) and the US NSA developed the Data Encryption Standard (DES) in 1977
    - DES is no longer strong enough (has been cracked)
    - http://www.networkworld.com/news/1999/0120cracked.html

DES Cracking Contest

- Operated by RSA
  - The distributed.net team, a non-profit organization of computer hobbyists, deciphered a DES-encoded message in just over 22 hours as part of the RSA Data Security's DES Challenge III
  - With an estimated 100,000 computers in the distributed.net (Uses idle times on the networked computers)
  - Just work hard… 245 billion keys per second; covered 22.2% of the available keyspace - some 72 quadrillion keys.

Next Generation – AES

- The so-called Advanced Encryption Standard is the replacement for the DES
  - Also a NIST standard (U.S. FIPS PUB 197)
  - U.S. Government: *The design and strength of all key lengths of the AES algorithm (i.e., 128, 192 and 256) are sufficient to protect classified information up to the SECRET level. TOP SECRET information will require use of either the 192 or 256 key lengths. The implementation of AES in products intended to protect national security systems and/or information must be reviewed and certified by NSA prior to their acquisition and use*

AES and the Future

- 128 bit keys implies 2^{128} (10^{39}) potential keys
  - Makes brute force attacks difficult
    - The current Storm botnet driving the enormous SPAM volumes in the past few weeks is estimated to control multiple million computers.
- *"80-bit key should offer an acceptable level of security for another 10 or 15 years"* (RSA)
- Absent a major breakthrough in quantum computing, it is unlikely that 128-bit keys will be broken by exhaustive search in the foreseeable future
- Concerns exist that alternative (clever) approaches might yield successful attacks.
  - XSL
  - Side channel
  - Timing

Block Cipher

- AES is a block cipher
  - Encryption transforms a fixed-length block of *plaintext* (unencrypted text) data into a block of *ciphertext*
  - Decryption reverses this process using the secret key
  - For us that means 8-bytes are encrypted at a time.
    - For single column encryption, that implies output columns are multiples of 8 bytes.

Modes of AES Encryption

- There are 4 types of block ciphers
  - *Cipher block chaining* – each plaintext block is XORed with the previous ciphertext block and then encrypted
  - *Electronic code book* – each plaintext block is encrypted independently with the block cipher (plaintext patterns are not concealed)
  - *Cipher feedback* – the previous ciphertext block is encrypted and the output produced is combined with the plaintext block using XOR to produce the current ciphertext block
  - *Output feedback* – similar to CFB mode except that the quantity XORed with each plaintext block is generated independently of both the plaintext and ciphertext

Business Response

- Encrypt all media (backup tapes) that leave the computer room.
  - Addresses the issue of media loss
  - Requires some method for key preservation
    - If the encryption key is “lost” the data is unrecoverable
    - Keys escrowed
    - Does not address employee theft of data
    - I recently received a letter documenting an incident where an employee of a vendor to my bank had sold my checking account information

Key Escrow Procedure

- Obviously, loss of the encryption keys would be a disaster
- Accordingly keys are escrowed in multiple ways in multiple locations
  - In three locations paper and machine readable (USB "thumb drives") format. These are in locked fireproof vaults.
    - Limited access to these vaults
  - In the DR site, a similar machine also contains the same key definitions on boot.
  - All three sites encrypt VMS backups
- Have yet to set a frequency for refreshment of the machine readable escrowed data.
  - Certainly once per year or two.

Business Response

- No longer display SSN information on screens and forms (generally only show last 4 digits)
  - Except for authorized individuals
  - And some reports (e.g. to collection agencies)
  - Does not address issues of people with ODBC access to the database
    - 9 tables have some SSN information for one reason or another.
    - After encrypting, the tables will have to be defragmented

Structure of an SSN

- SSN’s have three fields
  - 3-digit area number is assigned to geographical regions (today based on the zip (postal) code of the mailing address
    - List is available at http://www.socialsecurity.gov/employer/stateweb.htm
    - Large ranges of area numbers are unassigned
  - 2-digit group number
    - The Social Security Administration publishes the last group number used for each area number.
    - Since group numbers are allocated in a regular (if unusual) pattern, it is possible to identify an unissued SSN that contains an invalid group number
  - 4-digit serial number
- Some individuals do use duplicate SSNs

Employer ID Number

- An EIN or Federal tax number is used to identify business entities
- Corporate equivalent of SSN
- First field is the area where the EIN was assigned. It is 2 digits long
  - Not all area numbers are used
- Last 7 digits are a serial number

Storing SSN and EIN

- Usually SSN and EIN are stored in the same column in the database
- SSN is often not reported or given and is therefore stored as blank.
- All SSN/EIN are numeric

Difficulties Encrypting SSN

- SSN columns have limited range and structure
- One-third of the rows have no data [spaces]
- Have limited size [9 characters, default blank]

- Accordingly they represent a reasonably easy target for any decryption attack

DBA Response

- Encrypt all SSN data in the database
  - Target will be 16 bytes
  - Augment source with random characters to fill out to 16 source bytes
  - Considering randomizing the unused bits in the character field.
  - Special case the “no value” value and do not encrypt
- Provide stored procedures that perform
  - Encryption (used to store a value)
  - Decryption (requires privilege to execute)
  - Obfuscation (decrypt and x-out all but last 4 digits)
- Use external functions to encapsulate the algorithms

Caution

- Encryption is only as good as:
  - The secrecy of the secret key
  - The algorithm itself
  - The community's lack of knowledge on how to break keys
- No encryption can be expected to last forever
- The goal should be to make
  - the cost of breaking higher than it is worth
  - Strong enough to endure for a long period

VMS 8.3 Encryption Routines

- VMS 8.3 incorporates a full set of encryption routines that implement both DES and AES standard
  - DES is legacy and should not be used for new projects
  - The AES implementation supports all four modes
  - The AES supports 128, 192 and 256 bit algorithms
  - Default AES is 128-bit cipher block chaining – which was selected for this project

VMS Encryption Routines

There are two sets of routines:

- Encrypt/Decrypt a single record (column)
  - Requires initialization and teardown of the encryption context for each instance
  - Used for the data in this talk
- Encrypt/Decrypt large structures such as files etc.
  - Requires persistent context
  - Considering for final release

Encryption Keys

- Access to the VMS server is extremely limited
- Only to trusted individuals
- Encryption keys are defined at system boot time in the system encryption table
  - They are themselves encrypted with a VMS key
- Different keys are used for each encryption function:
  - VMS backup
  - Database
  - Column(s)

Encryption Keys

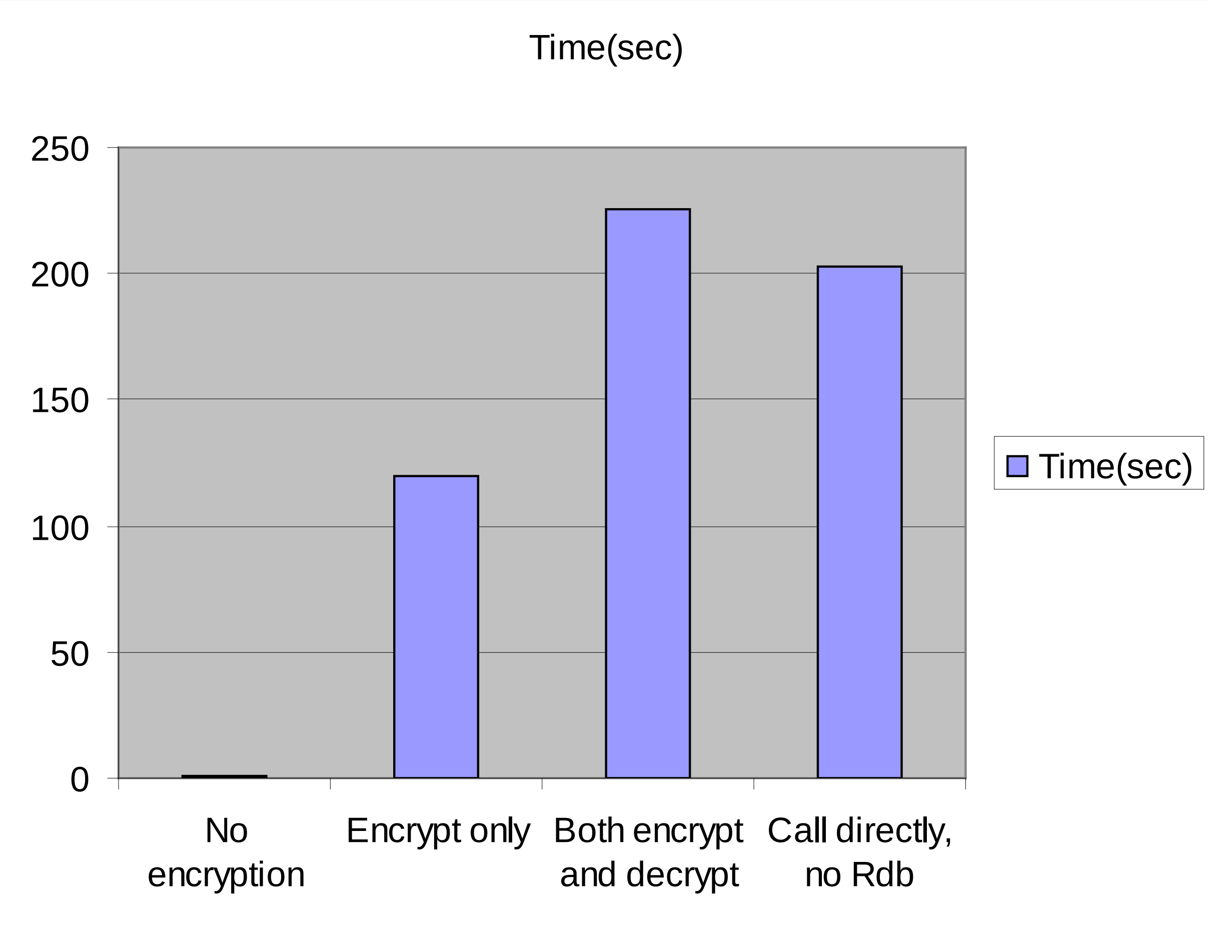
- Key string created as a long mixed-case alphanumeric string
  - Generated by the VMS command “set password/generate”
- Numeric characters interspersed and some letters uppercased

```
$ encrypt /create_key SAMPLE_KEY                                            -
    "baChGa2lle worOmet4orCh 7prothrohesS cirl6zbELe aUsti1oner"    -
    / aes/system
```

Nobody Encrypts for Performance

- We performed several tests to measure the performance of encryption
  - Loop through a single table with SSN column
  - About 287,319 rows
  - Various tests
    - Do nothing – measures the performance of the loop (1.19 sec)
    - Encrypt only – measures encryption performance (119.4 sec)
    - Encrypt and decrypt (225.43 sec)
  - A 4th test was run which called encrypt and decrypt 287,319 times (203 sec)
    - Yields insight into the overhead Rdb experiences in calling external functions

Timing for the Tests



Nobody Encrypts for Convenience

- Managing an encrypted environment adds management overhead
  - Forever
- Should one want to change the encryption keys, all data would have to first be decrypted and encrypted again with the new key.

Conclusions

- Encryption of single columns with limited range of values can be tricky because data patterns may be used for a non-random attack
- Encryption is a costly (CPU) process. Decryption is slightly less costly
  - Implies investigation of methodologies to retain context
- The mechanisms to accomplish this are straightforward.
- Key escrow is essential otherwise the data is lost

Questions

http://www.jcc.com

Join the worldwide Rdb community. Send mail to

OracleRdb-request@JCC.com with “SUBSCRIBE” in the body of the message.

Send additional questions to: Info@JCC.com